

# CONJUGATE VARIABLES AS A RESOURCE IN SIGNAL AND IMAGE PROCESSING

MICHAEL NÖLLE AND MARTIN SUDA

AIT–AUSTRIAN INSTITUTE OF TECHNOLOGY  
SAFETY & SECURITY DEPARTMENT  
OPTICAL QUANTUM TECHNOLOGIES

**ABSTRACT.** In this paper we develop a new technique to model joint distributions of signals. Our technique is based on quantum mechanical conjugate variables. We show that the transition probability of quantum states leads to a distance function on the signals. This distance function obeys the triangle inequality on all quantum states and becomes a metric on pure quantum states. Treating signals as conjugate variables allows us to create a new approach to segment them.

**Keywords:** Quantum information, transition probability, Euclidean distance, Fubini-study metric, Bhattacharyya coefficients, conjugate variable, signal/sensor fusion, signal and image segmentation.

## 1. INTRODUCTION

Quantum information theory and application areas based upon it, such as quantum computing or quantum cryptography have attracted a massive research interest and are developing rapidly. Quantum computing and quantum cryptography are primarily concerned with the impact that the nature of physical quantum systems has on the respective fields of computing and cryptography. Using physical systems, one has to obey the restrictions imposed by nature on physical quantum states, for example that they are not directly observable. They can be measured, but in general a measurement only reveals parts of the information contained in the quantum state.

Some authors have applied quantum information theory to statistics and probability theory ([3, 4]) as well as signal and image processing applications ([5, 8]). Their work focuses mainly on the abstract mathematical concept of quantum information theory and might be referred to as Quantum Information processing Algorithms (QIA) in analogy to the term Quantum Signal Processing created by Eldar ([8]). Applying the formalism of quantum information theory to information processing on classical computers can result in novel but still 'classical' algorithms.

In this paper we follow the QIA approach, motivated by the fact that we have found considerably easier descriptions of well suited algorithmic solutions for the

---

We would like to thank Jozef Gruska and Matthias Jakob for all the valuable discussions as well as Ian Glendinning and Janice Knight for editing the manuscript and the discussions thereabout. We also want to thank our company, the AIT, that enabled this research and the tax payers in Austria who paid for it. Last but not least we thank the reviewers who declined our project proposals on the subject and those who fostered it (to no avail) as they prompted us to show the value in this research even harder.

problems at hand in several application areas ([20, 23, 21]) within the mathematical setting of the quantum space. One reason might be that we were forced to think more deeply about the normalization of data and observations and the metrical functions to be used. A crucial factor seems to be the ability to represent some relations between two (or more) sources of information jointly as quantum information, i.e. in a quantum state. In analogy to the behaviour of physical conjugate variables, e.g. position and momentum, which play a fundamental role in quantum mechanics, we term this a *conjugate information variable*. In essence we refer to two sources of information as being conjugate to each other if they are in a special relation: Whenever one of them becomes highly predictable the other one is either undefined or unpredictable and vice versa.

The special relation between physical observations of conjugate quantities was an important factor in the development of quantum mechanics in the early 20<sup>th</sup> century. In the quantum mechanical setting, Heisenberg's uncertainty principle expresses this relation, e.g. between momentum and position. To be more precise, momentum and position are represented as operators in quantum mechanics and they are related to each other by derivatives. Here an exact definition of uncertainty means that these operators do not commute.

Our main justification for using the same mathematical structure, i.e. Hilbert spaces, is that information sources exist, which have a similar conjugate relation. As an illustrative example consider we want to observe two species of birds let's say one is blue the other one green and their chirp is quite distinct but cannot be reliably heard during day time due to background noises which are absent at night. Obviously during the day we use our eyes and at night our ears to distinguish the species. But is there a seamless way to fuse both signals that additionally helps us to distinguish them at sunrise or sunset with higher reliability?

Or more seriously, consider a simple classification task where we want to decide whether parts of a function are locally constant, rising, falling or minimum or maximum. We might want to use the derivatives of the function to achieve this, as we know that when the first derivative becomes zero the entire information on the extreme points of the function relies on the sign of the second (or third...) derivative. On the other hand when the second derivative becomes zero, the information whether we found a point of (rising or falling) inflection or the function is constant at this point depends entirely on the first (or third...) derivative. Again it would be desirable to represent the first and second derivative such that in 'grey areas' in between the two extremes either one alone (if it is more reliable) or both may contribute to the classification result.

The authors are fully aware that most of the material on theoretical quantum information is already covered elsewhere. In particular we refer to the textbooks of Gruska [12] and Nielsen and Chuang [19]. Nonetheless we find it worthwhile to summarize some of the basic ideas behind quantum information theory in order to help readers who are not too familiar with them and the notation used in this area. We will omit reoccurring references to [12, 19], where most of the material is treated in-depth, and indicate when the result can be found elsewhere. This introduction to quantum information theory is contained in Chapter 2 whereas Chapter 3 analyzes the encoding of information into quantum states. In Chapter 4 we take a closer look on two-dimensional systems. A new approach to signal segmentation is given

in Chapter 5. We conclude the paper with some discussion and references to future work (Chapters 6, 7).

The fact that in most parts of this paper we consider random variables and distributions rather than signals reflects the fact that we assume the signals to be normalized. In quantum states/signals the same assumption is usually made.

## 2. AN INTRODUCTION TO QUANTUM INFORMATION THEORY

**2.1. Quantum Systems, Quantum States and Qubits.** The first postulate of quantum mechanics states that any physical system can be described by a unit vector in an associated complex vector space which is called Hilbert space  $\mathcal{H}$ . A given unit vector  $\vec{\psi}$  is called the state of the system. In this paper we restrict ourselves to finite dimensional Hilbert spaces. The simplest non-trivial system is a two-dimensional system with the state space  $\mathcal{H} = \mathbb{C}^2$ . Such systems are called *quantum bits* or *qubits*. In the usual mathematical notation, the state of a qubit can thus be written as the vector

$$(2.1) \quad \vec{\psi} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha \hat{e}_0 + \beta \hat{e}_1 \quad \text{with} \quad \alpha, \beta \in \mathbb{C} \quad \text{and} \quad |\alpha|^2 + |\beta|^2 = 1,$$

where the basis vectors  $\hat{e}_0$  and  $\hat{e}_1$  are orthonormal unit vectors. In quantum computing, it is customary to use the Dirac (or “bra-ket”) notation: The column “ket”-vectors

$$(2.2) \quad |0\rangle \equiv \hat{e}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |1\rangle \equiv \hat{e}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

or generally  $|i\rangle \equiv \hat{e}_i$  for higher dimensional cases, form the *computational* or canonical basis of  $\mathcal{H}$  and the above state can be written as a *superposition*

$$(2.3) \quad |\psi\rangle = \alpha |0\rangle + \beta |1\rangle.$$

The *dual* “bra”-vector  $\langle\psi|$  corresponds to the associated row-vector  $(\alpha^*, \beta^*)$  with complex conjugate components, thus

$$(2.4) \quad \langle\psi| = \alpha^* \langle 0| + \beta^* \langle 1|.$$

The scalar product  $\langle\psi|\chi\rangle$  and the outer product  $|\psi\rangle\langle\chi|$  are thus reduced to matrix multiplications. Any linear operator  $\mathbf{A} : \mathcal{H} \rightarrow \mathcal{H}$  over  $\mathcal{H} = \mathbb{C}^n$  can be written as

$$(2.5) \quad \mathbf{A} = \sum_{i,j}^{n-1} a_{i,j} |i\rangle \langle j|.$$

For a *Hermitian operator*  $\mathbf{A} = \mathbf{A}^\dagger$  holds, where in matrix notation  $\mathbf{A}^\dagger$  is the transpose of  $\mathbf{A}$  with conjugate complex entries. Whenever we omit the index bounds, the sum is from  $0, \dots, n-1$ , where  $n$  is the dimension of the considered Hilbert space.

The orthonormality and completeness of the computational basis can be written as

$$(2.6) \quad \langle i|j\rangle = \delta_{ij} \quad \text{and} \quad \sum_i |i\rangle \langle i| = \mathbb{I},$$

where  $\delta_{ij}$  denotes the Kronecker delta function and  $\mathbb{I}$  is the identity. Table 1 summarizes frequently-used Dirac notation and its meaning.

Notation	Description
$ \psi\rangle$	general “ket” vector, e.g. $ \psi\rangle = (c_0, c_1, \dots)^T$
$\langle\psi $	dual “bra” vector to $ \psi\rangle$ , e.g. $\langle\psi  = (c_0^*, c_1^*, \dots)$
$ n\rangle$	$n^{\text{th}}$ basis vector of computational basis $N = \{ 0\rangle,  1\rangle, \dots\}$
$\langle\phi \psi\rangle$	inner product of $ \phi\rangle$ and $ \psi\rangle$ (scalar)
$ \phi\rangle\langle\psi $	outer product of $ \phi\rangle$ and $ \psi\rangle$ (matrix)
$ \phi\rangle \psi\rangle$	tensor product $ \phi\rangle \otimes  \psi\rangle$ (vector)
$ i, j\rangle$	tensor product $ i\rangle j\rangle$ of the basis vectors $ i\rangle$ and $ j\rangle$

TABLE 1. *Dirac (“bra-ket”) Notation*

**2.2. Measurement, von Neumann Measurement.** From a physical point of view any information contained in a quantum state  $|\psi\rangle$  can only be accessed via a *measurement* which is described by a set of (hermitian) operators adding up to the identity operator (completeness condition):

$$(2.7) \quad \sum_{i=0}^{m-1} \mathbf{M}_i = \mathbb{I}.$$

As a result of a measurement on  $|\psi\rangle$  we get an index  $i \in \{0, \dots, m-1\} = S$  with some probability. A standard interpretation of a measurement is that the state  $|\psi\rangle$  ‘collapses’ onto the post-measurement state  $\frac{1}{\sqrt{p(i)}} M_i |\psi\rangle$  with probability  $p(i) = \langle\psi| M_i |\psi\rangle$ . It should be noted that the idea that the quantum state spontaneously is changed via a measurement has been the subject of much debate ever since it was proposed in the 1950s. We do not want to address this issue, as the material discussed here is unaffected by it. The fact that the only direct information we can gain from a physical quantum system is the outcome  $i$  of the measurement is undisputed. Reapplying the measurement does not change this result any further. This means in particular that we cannot gain any direct information about the co-ordinates of a quantum state.

However if we are given a large number of ‘identical’ quantum states and we measure them one by one, we eventually gain the frequency of each index and after proper normalization the probability information  $p(i)$  for all  $i \in S$ . Taking the completeness condition (2.7) into account we see that the probabilities properly add up to one:  $\langle\psi| (\sum_i^m \mathbf{M}_i) |\psi\rangle = \langle\psi| \mathbb{I} |\psi\rangle = 1$ . In slightly different terms we get the probability distribution of a discrete random variable over the given index set  $P(X = i) = p(i)$ . The values the random variable can take<sup>1</sup> and the index of the measurement operators are in a one to one correspondence. The resulting probability distribution is fixed as soon as we are given the state  $|\psi\rangle$  and the set of measurement operators  $\mathbf{M} = \{\mathbf{M}_{i=0}^{m-1}\}$ . In the next section we will see that the converse in general is not true. That is, a given probability distribution does not uniquely determine a quantum state and/or a set of measurement operators.

In this paper we will restrict ourselves to a particularly simple form of measurement which is described by projectors, the so-called von Neumann measurement.

---

<sup>1</sup>the elements of the  $\sigma$  algebra.

Given a unit vector  $|\psi\rangle$  the projector  $\mathbf{P}$  onto  $|\psi\rangle$  takes the form

$$(2.8) \quad \mathbf{P} = |\psi\rangle \langle\psi|.$$

Applying  $\mathbf{P}|\psi\rangle = |\psi\rangle \langle\psi|\psi\rangle = 1 \cdot |\psi\rangle$  we see that it projects  $|\psi\rangle$  onto itself and any other unit vector  $|\chi\rangle$  is projected onto:  $|\psi\rangle \langle\psi|\chi\rangle = \langle\psi|\chi\rangle |\psi\rangle$ , i.e. in the direction of  $|\psi\rangle$  times the scalar  $\langle\psi|\chi\rangle$ . If  $\mathbf{P}^\perp = \mathbb{I} - \mathbf{P}$  is the projector onto the orthogonal complement space of  $\mathbf{P}$ ,  $\{\mathbf{P}, \mathbf{P}^\perp\}$  forms a measurement. For any state  $|\chi\rangle$  the probability that it is changed into  $|\psi\rangle$  via this measurement is

$$(2.9) \quad p(|\chi\rangle \rightarrow |\psi\rangle) \equiv \langle\chi|\mathbf{P}|\chi\rangle = \langle\chi|\psi\rangle \langle\psi|\chi\rangle = |\langle\psi|\chi\rangle|^2,$$

whereas

$$(2.10) \quad \langle\chi|\mathbf{P}^\perp|\chi\rangle = \langle\chi|(\mathbb{I} - \mathbf{P})|\chi\rangle = 1 - |\langle\psi|\chi\rangle|^2$$

gives the probability that the post measurement state lies in the orthogonal complement space.  $p(|\chi\rangle \rightarrow |\psi\rangle) = |\langle\psi|\chi\rangle|^2$  is called the transition probability. In the next section we will analyze these equations further.

**2.3. Operator Space, Density Operator.** The quantum state space under consideration is usually called the principal system. The *operator space* over a principal system is given by all linear mappings of this state space onto itself, i.e. given the state space  $\mathcal{H}$  the operator space is  $\mathcal{H} \otimes \mathcal{H} = L_{\mathcal{H}}$ . The inner product in the operator space takes the form of the Hilbert-Schmidt inner product:

$$(2.11) \quad (\mathbf{A}; \mathbf{B}) = \text{tr}(\mathbf{A}^\dagger \mathbf{B}); \quad \mathbf{A}, \mathbf{B} \in \mathcal{H} \otimes \mathcal{H},$$

where the trace of an operator is the sum of the diagonal elements of the operator written as a matrix. In terms of the Dirac notation the trace of an operator may be written as

$$(2.12) \quad \begin{aligned} \text{tr}(\mathbf{A}) &= \text{tr} \left( \sum_{i,j} a_{i,j} |i\rangle \langle j| \right) = \sum_k \langle k| \left( \sum_{i,j} a_{i,j} |i\rangle \langle j| \right) |k\rangle \\ &= \sum_k^n a_{k,k}. \end{aligned}$$

Equation 2.11 enables us to define a *norm* on  $L_{\mathcal{H}}$ :

$$(2.13) \quad \begin{aligned} \|\mathbf{A}\|^2 &= \text{tr}(\mathbf{A}^\dagger \mathbf{A}) = \text{tr} \left( \left( \sum_{i,j} a_{i,j} |i\rangle \langle j| \right)^\dagger \sum_{l,m} a_{l,m} |l\rangle \langle m| \right) \\ &= \text{tr} \left( \sum_{i,j} a_{i,j}^* |j\rangle \langle i| \sum_{l,m} a_{l,m} |l\rangle \langle m| \right) \\ &= \text{tr} \left( \sum_{i,j,m} a_{i,j}^* a_{i,m} |j\rangle \langle m| \right) = \sum_{i,j} a_{i,j}^* a_{i,j}. \\ \|\mathbf{A}\| &= \sqrt{\sum_{i,j} a_{i,j}^* a_{i,j}} = \sqrt{\sum_{i,j} |a_{i,j}|^2}. \end{aligned}$$

(2.13) is called the *Frobenius* norm or *Hilbert-Schmidt* norm of  $L_{\mathcal{H}}$ . Consequently, for operators  $\mathbf{A}, \mathbf{B} \in L_{\mathcal{H}}$  we may calculate the *Euclidian distance*  $\|\mathbf{A} - \mathbf{B}\|$  as:

$$\begin{aligned} \|\mathbf{A} - \mathbf{B}\| &= \sqrt{\sum_{i,j} (a_{i,j}^* - b_{i,j}^*)(a_{i,j} - b_{i,j})} = \sqrt{\sum_{i,j} [|a_{i,j}|^2 + |b_{i,j}|^2 - 2\Re(a_{i,j} b_{i,j}^*)]} \\ (2.14) \quad &= \sqrt{\|\mathbf{A}\|^2 + \|\mathbf{B}\|^2 - 2\Re \text{tr}(\mathbf{A}\mathbf{B}^\dagger)}, \end{aligned}$$

where  $\Re$  gives the real part of a complex number.

Finally we would like to emphasis the well known fact that global phase changes of a state do not affect the expectation value of an operator.

**Lemma 2.1** (Global phase invariance of the expectation of linear operators). *Let  $A \in L_{\mathcal{H}}$  be a linear operator. Then the expectation of  $A$  is invariant with respect to global phase changes, i.e. for any  $|\psi\rangle = e^{i\varphi} |\psi\rangle$ ,  $-\pi \leq \varphi \leq \pi$ ,  $|\psi\rangle \in \mathbb{C}^n$ :*

$$\langle \tilde{\psi} | \mathbf{A} | \tilde{\psi} \rangle = \langle \psi | \mathbf{A} | \psi \rangle.$$

This is easily verified by

$$\textit{Proof.} \quad \langle \tilde{\psi} | \mathbf{A} | \tilde{\psi} \rangle = e^{-i\varphi} \langle \psi | \mathbf{A} e^{i\varphi} | \psi \rangle = e^{-i\varphi+i\varphi} \langle \psi | \mathbf{A} | \psi \rangle = \langle \psi | \mathbf{A} | \psi \rangle. \quad \square$$

After these general remarks on the operator space we will turn our attention to a special subset of operators, the so called *density operators*. Lets first consider the single qubit state

$$(2.15) \quad |+\rangle = \frac{1}{\sqrt{2}} |0\rangle + \frac{1}{\sqrt{2}} |1\rangle.$$

When we measure  $|+\rangle$  by projectors onto the computational basis  $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ , we will find that either of the two possible post-measurement states  $|0\rangle$  and  $|1\rangle$  are equally likely with a transition probability of  $p_0 = p_1 = \frac{1}{2}$ . So the experiment is equivalent to a classical flip of an unbiased coin and we might be tempted to think that  $|+\rangle$  in fact describes a state where the system state is *unknown* which is either  $|0\rangle$  or  $|1\rangle$  with equal probability.

If, however, we measure via  $\{|+\rangle\langle +|, |-\rangle\langle -|\}$  with

$$(2.16) \quad |-\rangle = \frac{1}{\sqrt{2}} |0\rangle - \frac{1}{\sqrt{2}} |1\rangle,$$

on the same state  $|+\rangle$ , then the outcome is deterministic as  $|+\rangle$  is an eigenvector of  $|+\rangle\langle +|$ . Obviously, the non-determinism of a quantum measurement is inherently different from the classical randomness of (un)biased coin flips.

There exists an elegant and consistent way to capture both concepts of randomness: Assuming we know that a quantum system is in one of  $k$  states  $|\psi_i\rangle$  with the probabilities  $p_i$ , then we can aggregate our knowledge of the state in a (semi) positive *density operator*

$$(2.17) \quad \rho = \sum_{i=1}^k p(i) |\psi_i\rangle \langle \psi_i| \quad \text{with} \quad \sum_{i=1}^k p(i) = 1.$$

As the probabilities add up to 1 and by the linearity of the trace, it is assured that  $\text{tr} \rho = \text{tr} \left( \sum_{i=1}^k p_i |\psi_i\rangle \langle \psi_i| \right) = \sum_{i=1}^k p_i \text{tr} |\psi_i\rangle \langle \psi_i| = \sum_{i=1}^k p_i = 1$  where we used Eq. 2.12. If  $\text{tr} \rho^2 = 1$  then  $\rho$  can be written in the form  $\rho = |\psi\rangle \langle \psi|$  and the

system is said to be in a *pure state*. If  $\text{tr } \rho^2 < 1$ , then  $\rho$  is referred to a *mixed state*. The transition probability  $p(\rho \rightarrow \chi)$  to reduce  $\rho$  (pure or mixed state) to the pure state  $\rho' = |\chi\rangle\langle\chi|$  when performing a measurement in a basis containing  $|\chi\rangle$  is the *expectation value*

$$(2.18) \quad p(\rho \rightarrow \chi) = \langle \rho \rangle_\chi = \langle \chi | \rho | \chi \rangle.$$

Note that for pure states, (2.18) is equivalent to (2.9), as we would expect.

There are some properties of density operators that should be mentioned:

- $\text{tr } \rho = 1$  (see above);
- $\rho = \rho^\dagger$ , i.e. density operators are hermitian operators;
- any density operator has a spectral decomposition,  $\rho = \mathbf{U} \mathbf{D} \mathbf{U}^\dagger$ , where  $\mathbf{U}$  is a unitary matrix, i.e.  $\mathbf{U}^\dagger \mathbf{U} = \mathbb{I}$ , and  $\mathbf{D}$  is a diagonal matrix with nonnegative real diagonal elements. Therefore,  $\rho = \sum_i d(i) |u_i\rangle\langle u_i|$ , with  $\sum_i d(i) = 1$ , where  $|u_i\rangle$  is the basis generated by the columns of  $\mathbf{U}$  and  $d_i$  are the diagonal entries of  $\mathbf{D}$ . The  $|u_i\rangle$  are called the eigenbasis of  $\rho$ ;
- consider two density operators in their eigenbasis representation:  
 $\rho = \sum_i p(i) |\psi_i\rangle\langle\psi_i|$  with  $\sum_i p(i) = 1$  and  $\sigma = \sum_i q(i) |\chi_i\rangle\langle\chi_i|$  with  $\sum_i q(i) = 1$ , then  $\text{tr } (\rho\sigma) = \sum_{i,j} p(i)q(j) |\langle\psi_i|\chi_j\rangle|^2$ . This implies, as  $p_i, q_j$  and  $|\langle\psi_i|\chi_j\rangle|^2$  are nonnegative and real numbers, that  $\text{tr } (\rho\sigma) = \Re \text{tr } (\rho\sigma)$ ;
- $0 \leq \text{tr } (\rho\sigma) \leq 1$ . The lower bound follows again from the fact that the sum is over nonnegative and real numbers. To verify the upper bound we substitute  $|\langle\psi_i|\chi_j\rangle|^2 \leq 1$ , which gives

$$(2.19) \quad 0 \leq \text{tr } (\rho\sigma) = \sum_{i,j} p(i)q(j) |\langle\psi_i|\chi_j\rangle|^2 \leq \sum_{i,j} p(i)q(j) = 1.$$

The minimum error discrimination of mixed quantum states is an interesting and well studied problem. We refer the interested reader to pioneering work of [13, 15] and further results in [9, 14] and the literature cited therein.

### 3. ENCODING CLASSICAL INFORMATION AS QUANTUM INFORMATION

In Section 2.2 we have seen that given a set of measurement operators  $\mathbf{M} = \{\mathbf{M}_{i=0}^{m-1}\}$  and a quantum system  $|\psi\rangle$  uniquely determines a probability distribution  $P(X=i) = p(i), i=0, \dots, m-1$ . In this section we want to analyze the converse, i.e. given a probability distribution we want to determine quantum states and measurement operators that produce it.

Let us consider a random variable  $X$ . The set of values  $X$  can take is given by  $S$  and for any probability distribution

$$(3.1) \quad \sum_{x \in S} p(X=x) = 1, \quad p(X=x) \geq 0, \quad \text{for all } x \in S.$$

A given probability distribution may be interpreted as vector from the  $n = |S|$  dimensional real vector space  $\mathbb{R}^n$ . In this case the basis of the space can be indexed by the elements of  $S$  and  $|x\rangle, x \in S$  denotes a basis vector. In this basis we define  $\overrightarrow{p(X)} \equiv |p(X)\rangle = \sum_{x \in S} p(x) |x\rangle$ . Due to the linearity of the normalization condition (3.1) the set of all probability distributions forms a  $n-1$  dimensional hyper-plane, sometimes called simplex, in this space.

As we would like to make a connection between the set of all probability distributions in  $\mathbb{R}^n$  and quantum states from a complex Hilbert space  $\mathbb{C}^n = \mathcal{H}$ , we

formerly introduce complex numbers  $\alpha_{\varphi_x} \in \mathbb{C}$  such that

$$(3.2) \quad p(x) = \alpha_{\varphi_x}^* \alpha_{\varphi_x} = |\alpha_{\varphi_x}|^2.$$

The choice of  $\alpha_{\varphi_x}$  is not unique. Using Euler's Equation on

$$\alpha_{\varphi_x} = \sqrt{p(x)} e^{i\varphi_x} = \sqrt{p(x)} (\cos \varphi_x + i \sin \varphi_x), \quad -\pi \leq \varphi_x \leq \pi,$$

Equation (3.2) can be rewritten as

$$(3.3) \quad p(x) = \alpha_{\varphi_x}^* \alpha_{\varphi_x} = p(x) \cos^2 \varphi_x + p(x) \sin^2 \varphi_x.$$

It is noteworthy that all numbers  $\alpha_{\varphi_x}$  fulfilling (3.2) lie on a complex sphere with radius  $\sqrt{p(x)}$  as can be verified from (3.3).

Already here we see that encoding probabilities as quantum states requires some more information than only the probability distribution if we want to be able to adress all quantum states<sup>2</sup>. If we restrict the encoding onto the real nonnegative square root we gain the same quantum state as described in [11].

Consider the (state) vector  $|\psi\rangle \in \mathbb{C}^n$ :

$$(3.4) \quad |\psi\rangle = \sum_{x \in S} \alpha_{\varphi_x} |x\rangle.$$

The set of all projectors onto basis states,  $\{\mathbf{P}_{|x\rangle} = |x\rangle \langle x|, x \in S\}$ , taken as measurement operators result in the probability distribution  $p(X)$  when  $|\psi\rangle$  is measured, and fulfill the completeness condition (2.7).

Using (3.3) and the normalization condition (3.1) we see that  $|\psi\rangle$  is a unit vector in  $\mathbb{C}^n$ . These unit vectors are related to probability distributions via (3.2) and lie on a complex unit hyper-sphere centered at the origin.

We will call the assignment

$$(3.5) \quad E(p(X=x), C) \rightarrow \alpha_{\varphi_x}$$

quantum encoding of  $p(X)$  with respect to the conjugate information  $C$ . If we want to decode the probability information of the quantum encoding we have to apply  $\{\mathbf{P}_{|x\rangle} = |x\rangle \langle x|, x \in S\}$  to  $|\psi\rangle$ .

**3.1. Distances.** When given two probability distributions over the same index set a natural and important question is: how similar or how close are they? The answers to this question can be quite diverse depending on the scientific discipline they have been derived from. Table 2 gives an overview on the more frequently used probability distribution distance measures as they can be found in the literature.

Likewise we find adapted distance measures for quantum states and/or density operators ([9]). Based on the transition probability of quantum states we may define a distance function:

**Lemma 3.1.** *Let  $\rho, \sigma \in L_{\mathcal{H}}$  be density operators, ( $\text{tr } \rho = \text{tr } \sigma = 1$ ). Then the function*

$$D(\rho, \sigma) = \sqrt{1 - \text{tr } (\rho\sigma)}$$

*satisfies the following properties:*

$$(1) \quad 0 \leq D(\rho, \sigma) \leq 1,$$

---

<sup>2</sup>To be more precise: we need to know the meaning of ' $\sqrt{p(x)}$ '. This cannot be derived from the probability distribution alone. One bit of additional information is required. If this bit of information is available, the complex part insures a smooth transition from  $\sqrt{p(x)}$  to  $-\sqrt{p(x)}$ , i.e. without changing the probability of  $|x\rangle$  if measured in the same basis.



Name	Formula	Metric
Euclidian distance	$\sqrt{\sum_i (p(i) - q(i))^2}$	yes
trace or variational distance ([17])	$\sum_i  p(i) - q(i) $	yes
relative $\chi^2$	$\sum_i \frac{(p(i) - q(i))^2}{2(p(i) + q(i))}$	$\chi$ yes
relative entropy or Kullback-Leibler divergence ([16])	$\sum_i p(i) \log \frac{p(i)}{q(i)} = -H(p) - \sum_i p(i) \log q(i)$	no
Jensen-Shannon divergence or symmetric relative entropy ([17])	$JS(p, q) = H\left(\frac{1}{2}(p + q)\right) - \frac{1}{2}(H(p) + H(q))$	$\sqrt{JS(p, q)}$ yes
Bhattacharyya distance ([1, 16])	$1 - \sum_i \sqrt{p(i)q(i)}$	yes

TABLE 2. Frequently used probability distribution distance measures.  $p$  and  $q$  are probability distributions over the same index set.  $H(p) = -\sum_i p(i) \log p(i)$  is the Shannon entropy ([7, 10, 2]).

- (2)  $D(\rho, \sigma) = D(\sigma, \rho)$ , (symmetry)
- (3) for any density operator  $\tau \in L_{\mathcal{H}}$ ,  $\text{tr } \tau = 1$ :
 
$$D(\rho, \sigma) \leq D(\rho, \tau) + D(\tau, \sigma), \quad (\text{triangle inequality})$$
- (4) on pure states  $D(.,.)$  defines a metric.

The fact that  $D$  defines a metric on pure states is well known ([19], p. 500). In the literature it is sometimes called *no-name metric* and it is closely related to the so called Fubini study metric. Nevertheless, we would like to prove Lemma 3.1 as the proof offers some insight into the relation of  $D$  and the Euclidian metric of the operator space and the principal system.

*Proof.* Property (1) of Lemma 3.1 is a direct consequence of Equation (2.19). Property (2) follows from the fact that the trace of an operator product is invariant against cyclic permutations of the arguments, i.e.  $\text{tr}(\rho\sigma) = \text{tr}(\sigma\rho)$ .

Let  $\rho, \sigma, \tau \in L_{\mathcal{H}}$  be density operators, i.e.  $\text{tr } \rho = \text{tr } \sigma = \text{tr } \tau = 1$  and  $\|\rho\| \leq 1, \|\sigma\| \leq 1, \|\tau\| \leq 1$ . To see property (3) we first note that the Euclidian distance (see Equation 2.11) of the operator space fulfills the triangle inequality for all operators in  $L_{\mathcal{H}}$ , i.e.  $\|\rho - \sigma\| \leq \|\rho - \tau\| + \|\tau - \sigma\|$  holds. We calculate

$$\begin{aligned}
 \|\rho - \sigma\|^2 &\leq (\|\rho - \tau\| + \|\tau - \sigma\|)^2 \\
 \|\rho\|^2 + \|\sigma\|^2 - 2\text{tr}(\rho\sigma) &\leq \|\rho - \tau\|^2 + \|\tau - \sigma\|^2 + 2\|\rho - \tau\|\|\tau - \sigma\| \\
 &= \|\rho\|^2 + \|\tau\|^2 - 2\text{tr}(\rho\tau) + \|\tau\|^2 + \|\sigma\|^2 - 2\text{tr}(\tau\sigma) + \\
 &\quad + 2\|\rho - \tau\|\|\tau - \sigma\|,
 \end{aligned}$$

where we used Equation 2.14 in the last step. Simplifying and expanding the mixed term gives:

$$\begin{aligned}
 -\text{tr}(\rho\sigma) &\leq \|\tau\|^2 - \text{tr}(\rho\tau) - \text{tr}(\tau\sigma) + \\
 &\quad + \sqrt{[\|\rho\|^2 + \|\tau\|^2 - 2\text{tr}(\rho\tau)][\|\tau\|^2 + \|\sigma\|^2 - 2\text{tr}(\tau\sigma)]}.
 \end{aligned}$$

By substituting  $\|\rho\|^2 + \|\tau\|^2 \leq 2$  and  $\|\tau\|^2 + \|\sigma\|^2 \leq 2$  we get

$$\begin{aligned}
 (3.6) \quad -\text{tr}(\rho\sigma) &\leq \|\tau\|^2 - \text{tr}(\rho\tau) - \text{tr}(\tau\sigma) + \\
 &\quad + 2\sqrt{(1 - \text{tr}(\rho\tau))(1 - \text{tr}(\tau\sigma))}.
 \end{aligned}$$

With this we are ready to prove property (3) of the above Lemma. We substitute  $\|\tau\|^2 \leq 1$  in (3.6) and add 1 to both sides. Then

$$\begin{aligned} D(\rho, \sigma)^2 = 1 - \text{tr}(\rho\sigma) &\leq 2 - \text{tr}(\rho\tau) - \text{tr}(\tau\sigma) + \\ &+ 2\sqrt{(1 - \text{tr}(\rho\tau))(1 - \text{tr}(\tau\sigma))} \\ &= \left(\sqrt{1 - \text{tr}(\rho\tau)} + \sqrt{1 - \text{tr}(\tau\sigma)}\right)^2 \\ &= (D(\rho, \tau) + D(\tau, \sigma))^2. \end{aligned}$$

Reading top to bottom we see that the triangle inequality holds for  $D(.,.)$  and all  $\rho, \sigma, \tau \in L_{\mathcal{H}}$ :

$$(3.7) \quad D(\rho, \sigma) \leq D(\rho, \tau) + D(\tau, \sigma).$$

For property (4) we have to verify that  $D(\rho, \rho) = 0$ . This is only true for pure states as in general  $\text{tr} \rho^2 \leq 1$  with equality if and only if  $\rho$  is pure. On the other hand for pure states  $D(.,.)$  fulfills all the requirements of a metric.  $\square$

Indeed, for pure states  $D$  is identical to the Euclidian norm of the operator space up to a constant factor, which is easy to verify by noting that for any pure state  $|\psi\rangle$  and any density operator  $\sigma$  the following holds (see. Eq. 2.12):

$$(3.8) \quad \text{tr}(|\psi\rangle\langle\psi|\sigma) = \langle\psi|\sigma|\psi\rangle.$$

If  $\sigma$  is pure as well, i.e  $\sigma = |\xi\rangle\langle\xi|$ , we get  $\text{tr}(|\psi\rangle\langle\psi|\sigma) = \text{tr}(|\psi\rangle\langle\psi|\xi\rangle\langle\xi|) = |\langle\psi|\xi\rangle|^2$ . With this and equation 2.14 the distance computes to  $\| |\psi\rangle\langle\psi| - |\xi\rangle\langle\xi| \| = \sqrt{2}\sqrt{1 - |\langle\psi|\xi\rangle|^2}$  and  $D(|\psi\rangle\langle\psi|, |\xi\rangle\langle\xi|) = \sqrt{1 - |\langle\psi|\xi\rangle|^2}$ . For convenience we will write  $D(|\psi\rangle, |\xi\rangle) \equiv D(|\psi\rangle\langle\psi|, |\xi\rangle\langle\xi|)$  for short when dealing with pure states only.

On pure states the no-name metric takes therefore the form

$$(3.9) \quad D(|\psi\rangle, |\xi\rangle) = \sqrt{1 - |\langle\psi|\xi\rangle|^2}, \quad \text{with} \quad |\psi\rangle = \sum_x \alpha_{\varphi_x} |x\rangle, |\xi\rangle = \sum_x \beta_{\chi_x} |x\rangle.$$

One may derive an intuitive geometric argument to compare two state vectors in  $\mathbb{C}^n$  (see Section 3.2 and Figure 1). The vectors  $|\psi\rangle$  and  $|\xi\rangle$  in general span a two-dimensional complex plane and the intersection of this plane with the hyper-sphere generates a complex unit circle. If  $|\psi\rangle = |\xi\rangle$  the intersection degenerates into two points.

*Remark (a word of care):* Given two probability distributions of the same random variable  $S$  way may construct corresponding (diagonal) density operators by:  $\rho = \sum_{x \in S} p(x) |x\rangle\langle x|$ ,  $\sigma = \sum_{x \in S} q(x) |x\rangle\langle x|$ . Their Euclidian operator distance  $\|\rho - \sigma\|^2 = \text{tr}(\rho^2) + \text{tr}(\sigma^2) - 2\text{tr}(\rho\sigma)$  is equal to the Euclidian distance of the probability distributions. But  $\|\rho - \sigma\|^2 \neq 2(1 - \text{tr}(\rho\sigma)) = 2D(\rho, \sigma)^2$  unless both distributions are pure, i.e. for both probability distributions there exists one event, lets say  $x_p, x_q$ , such that  $p(x_p) = q(x_q) = 1$ .

For the rest of the chapter we will assume pure states only.

**3.1.1. Projection, transition probability.** Let  $|\psi\rangle = \sum_{x \in S} \alpha_{\varphi_x} |x\rangle$ ,  $\alpha_{\varphi_x} = \sqrt{p(x)}e^{i\varphi_x}$  and  $|\xi\rangle = \sum_{x \in S} \beta_{\chi_x} |x\rangle$ ,  $\beta_{\chi_x} = \sqrt{q(x)}e^{i\chi_x}$ . As we have seen in Section 2.2, Equation (2.8) the projector  $\mathbf{P}$  onto  $|\psi\rangle$  is given by  $\mathbf{P} = |\psi\rangle\langle\psi|$  and the transition probability

$$p(|\xi\rangle \rightarrow |\psi\rangle) = \langle\xi|\mathbf{P}|\xi\rangle = \langle\xi|\psi\rangle\langle\psi|\xi\rangle = |\langle\psi|\xi\rangle|^2$$

gives the square of the length of the projection of  $|\xi\rangle$  onto  $|\psi\rangle$ . Substituting  $\gamma_x = \varphi_x - \chi_x$  and using Euler's Equation, we get:

$$\begin{aligned} \langle \xi | \psi \rangle &= \sum_{x \in S} \sqrt{p(x)q(x)} e^{i\gamma_x} \\ (3.10) \quad &= \sum_{x \in S} \cos \gamma_x \sqrt{p(x)q(x)} + i \sum_{x \in S} \sin \gamma_x \sqrt{p(x)q(x)}, \end{aligned}$$

and finally

$$\begin{aligned} |\langle \xi | \psi \rangle|^2 &= \left( \sum_{x \in S} \cos \gamma_x \sqrt{p(x)q(x)} \right)^2 + \left( \sum_{x \in S} \sin \gamma_x \sqrt{p(x)q(x)} \right)^2 \\ &= \sum_{x, y \in S} \cos(\gamma_x - \gamma_y) \sqrt{p(x)p(y)q(x)q(y)}, \\ (3.11) \quad &= \sum_{x \in S} p(x)q(x) + \sum_{x, y \neq x \in S} \cos(\gamma_x - \gamma_y) \sqrt{p(x)p(y)q(x)q(y)}, \end{aligned}$$

where we used  $\cos(\gamma_x - \gamma_y) = \cos \gamma_x \cos \gamma_y + \sin \gamma_x \sin \gamma_y$  and note that all mixed terms  $(\pm i \cos \gamma_x \sin \gamma_y \dots)$  disappear.

**3.1.2. Length, Euclidian distance of the principal system.** It is useful to compute the Euclidian distance of the principal system as well in order to compare it with the metric  $D$ . This is given by the length of the difference  $\|\psi\rangle - |\xi\rangle$ :

$$\begin{aligned} \|\psi\rangle - |\xi\rangle\|^2 &= \sum_{x \in S} (\alpha_{\varphi_x} - \beta_{\chi_x})^* (\alpha_{\varphi_x} - \beta_{\chi_x}) \\ &= \sum_{x \in S} p(x) + q(x) - \sqrt{p(x)q(x)} (e^{i\gamma_x} + e^{-i\gamma_x}) \\ &= \sum_{x \in S} p(x) + q(x) - 2\sqrt{p(x)q(x)} \cos \gamma_x \end{aligned}$$

with  $\gamma_x = \varphi_x - \chi_x$ . Simplifying by using the normalization condition 3.1 and Equation 3.10 we get:

$$(3.12) \quad \|\psi\rangle - |\xi\rangle\|^2 = 2 - 2 \sum_{x \in S} \cos \gamma_x \sqrt{p(x)q(x)} = 2 (1 - \Re \langle \xi | \psi \rangle).$$

**3.1.3. Relation between Projection, no-name metric and Euclidian distance of the principal system.** Let  $\mathbf{P} = |\psi\rangle \langle \psi| = \mathbf{U}^\dagger \mathbf{D} \mathbf{U}$  be the spectral decomposition of  $\mathbf{P}$ , where  $\mathbf{D}$  is a diagonal matrix with non-negative diagonal entries, and  $\mathbf{U}$  is unitary. Clearly,  $\langle \psi | \mathbf{P} | \psi \rangle = 1$  and therefore,  $\mathbf{U} |\psi\rangle = |k\rangle$  is a basis vector in some appropriate basis system and  $\mathbf{D} = |k\rangle \langle k|$ . Define  $|\xi'\rangle = \mathbf{U} |\xi\rangle = \sum_j \alpha'_j |j\rangle$ . Then, with  $\alpha'_j = a_j + ib_j$ ,  $a, b \in \mathbb{R}$ ,

$$\langle \xi | \mathbf{P} | \xi \rangle = \langle \xi | \mathbf{U}^\dagger \mathbf{D} \mathbf{U} | \xi \rangle = \langle \xi' | k \rangle \langle k | \xi' \rangle = \alpha_k'^* \alpha'_k = a_k^2 + b_k^2,$$

or in terms of the no-name metric we get:

$$D(|\psi\rangle, |\xi\rangle)^2 = 1 - a_k^2 - b_k^2.$$

The Euclidian distance computes to (see 3.12):

$$\begin{aligned} \|\psi\rangle - |\xi\rangle\|^2 &= \|\mathbf{U} |\psi\rangle - \mathbf{U} |\xi\rangle\|^2 = \| |k\rangle - |\xi'\rangle \|^2 \\ &= 2 (1 - \Re \langle \xi' | k \rangle) = 2 (1 - \Re \alpha'_k) = 2 (1 - a_k). \end{aligned}$$

If we assume a real nonnegative principal system this might give some insight into distance functions used in the literature ([1, 22, 6]). In this case as  $b = 0, 0 \leq a \leq 1$  and  $1 - a \leq 1 - a^2$  we have (see Fig. 1):  $D(|\psi\rangle, |\xi\rangle) \leq \|\psi\rangle - |\xi\rangle\| \leq \sqrt{2}D(|\psi\rangle, |\xi\rangle)$ .

**3.2. Dropping the phase.** The relation between the Euclidian distance of the principal system and the no-name metric becomes slightly easier (but misleading in general) when we consider no phase information, i.e. when  $p$  and  $q$  are represented by their special nonnegative and real choice  $|\psi_p\rangle$  and  $|\xi_q\rangle$ . In this case the complex Hilbert space  $\mathbb{C}^n$  is reduced to a real Hilbert space of the same dimension and we may drop the modulus as all probabilities are mapped onto the positive square root. We can visualize the relation of the two measures as can be seen in Figure 1.

The inner product  $\langle \xi_q | \psi_p \rangle$  gives the cosine of the angle between  $|\psi_p\rangle$  and  $|\xi_q\rangle$  and equals the length of the projection of  $|\xi_q\rangle$  onto  $|\psi_p\rangle$ . Likewise  $\sqrt{1 - \langle \xi_q | \psi_p \rangle^2}$  results into the sine of the angle and equals the length of the projection of  $|\xi_q\rangle$  onto the orthogonal complement of  $|\psi_p\rangle$  in the plane spanned by the two state vectors. The length of the difference of the two vectors in terms of the inner product is given by (Eq. 3.12)

$$(3.13) \quad \|\psi_p\rangle - |\xi_q\rangle\| = \sqrt{2}\sqrt{1 - \langle \xi_q | \psi_p \rangle}.$$

This relates to a measure between density distributions frequently used in the literature that is based on the so called Bhattacharyya coefficients:

$$(3.14) \quad d(p(X), q(X)) \equiv d(p, q) = \sqrt{1 - \sum_{x \in S} \sqrt{p(x)q(x)}} = \sqrt{1 - \langle \xi_q | \psi_p \rangle},$$

which up to a constant factor is equal to (3.13). The sum over the Bhattacharyya coefficients,  $f(p(X), q(X)) = \sum_{x \in S} \sqrt{p(x)q(x)}$ , is usually called the *fidelity* of the probability distribution  $p$  and  $q$ . As an example 3.14 is used in [6] to compare density functions in the context of object tracking.

Let  $\omega$  be the angle between  $|\psi_p\rangle$  and  $|\xi_q\rangle$ . Substituting  $\cos \omega = \langle \xi_q | \psi_p \rangle$  and observing that  $\sin \frac{\omega}{2} = \frac{1}{\sqrt{2}}\sqrt{1 - \cos \omega}$ , we get:

$$\begin{aligned} D(p(X), q(X)) &= \sin \omega \\ \|\psi_p\rangle - |\xi_q\rangle\| &= \sqrt{2}d(p, q) = 2 \sin \frac{\omega}{2}. \end{aligned}$$

For the rest of the paper we will limit our discussion onto two dimensional systems as they allow to be visualized are fairly powerful instruments for concrete applications.

#### 4. VISUALIZATION OF $D$ , THE TWO-DIMENSIONAL CASE

It is worthwhile to study the behavior of the transition probability and the no-name metric more closely. For the two-dimensional Hilbert space  $\mathbb{C}^2$  and pure states  $|\psi\rangle, |\xi\rangle \in \mathbb{C}^2$  we can visualize Equation 3.11 and consequently  $D(|\psi\rangle, |\xi\rangle)$ .

Consider density distributions  $p, q$  over the random variable  $X = \{x, y\}$ . In this case the encoding of  $p$  and  $q$  as quantum states results in qubits. We want to analyze the impact of the phase factor on the distance of the qubits.

Using Equation 3.11 Equation (3.9) then takes the form

$$D(|\psi\rangle, |\xi\rangle)^2 = 1 - p(x)q(x) - p(y)q(y) - 2 \cos(\gamma_x - \gamma_y) \sqrt{p(x)p(y)q(x)q(y)}.$$

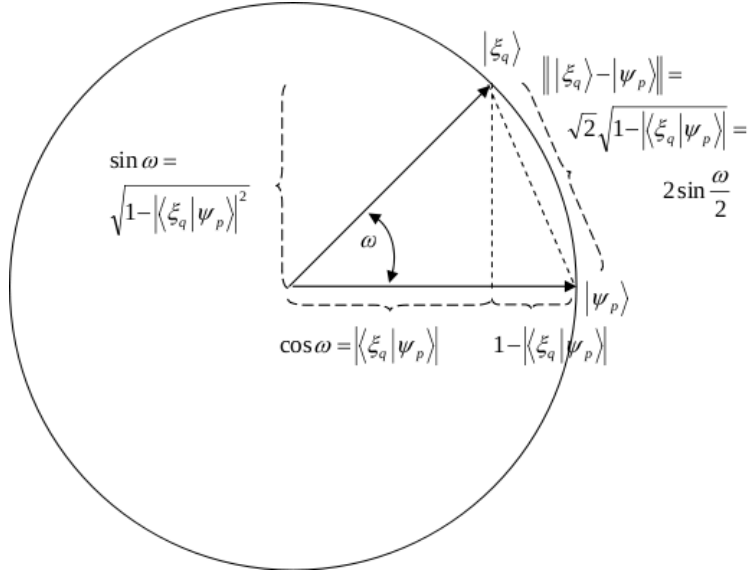


FIGURE 1. Geometric derivation of several distance measures in  $\mathbb{R}^n$ . The intersection of the plane spanned by  $|\psi_p\rangle$  and  $|\xi_q\rangle$  with the unit hypersphere of  $\mathbb{R}^n$  generates a unit circle.

Alternatively, we may express  $D(|\psi\rangle, |\xi\rangle)$  in terms of  $p(x) = |\alpha_x|^2$ ,  $p(y) = |\alpha_y|^2$ ,  $q(x) = |\beta_x|^2$  and  $q(y) = |\beta_y|^2$ :

$$D(|\psi\rangle, |\xi\rangle)^2 = 1 - |\alpha_x|^2 |\beta_x|^2 - |\alpha_y|^2 |\beta_y|^2 - 2 \cos(\gamma_x - \gamma_y) |\alpha_x| |\alpha_y| |\beta_x| |\beta_y|.$$

- (1) Let  $p$  be fixed and  $q$  vary over all possible distributions. As  $p$  is fixed the phase difference  $\gamma_x - \gamma_y$  only depends on  $q$  and we may consider all possible phase differences. A visualization of  $D(|\psi\rangle, |\xi\rangle)$  for several choices of  $p$  is given in Figure 2.

- (a) Clearly,  $D(|\psi\rangle, |\xi\rangle) = 0 \Leftrightarrow \alpha_x = \beta_x, \alpha_y = \beta_y$  as then  $\gamma_x - \gamma_y = 0$  and

$$D(|\psi\rangle, |\xi\rangle)^2 = 1 - (|\alpha_x|^2 + |\alpha_y|^2)^2 = 1 - (p(x) + p(y))^2 = 0.$$

- (b) Conversely,  $D(|\psi\rangle, |\xi\rangle) = 1 \Leftrightarrow \alpha_x = |\beta_y| e^{i(\chi_x \pm \frac{\pi}{2} + \eta)}$ ,

$$\alpha_y = |\beta_x| e^{i(\varphi_{q,y} \mp \frac{\pi}{2} + \eta)}, \eta \in \mathbb{R}, \text{ which gives } \gamma_x - \gamma_y = \pm\pi, \text{ resulting in}$$

$$D(|\psi\rangle, |\xi\rangle)^2 = 1 - |\beta_y|^2 |\beta_x|^2 - |\beta_x|^2 |\beta_y|^2 + 2 |\beta_x|^2 |\beta_y|^2 = 1.$$

- (2) Now let  $\gamma_x - \gamma_y$  be fixed and  $p, q$  vary over all possible distributions. A visualization of  $D(|\psi\rangle, |\xi\rangle)$  for several choices of  $\gamma_x - \gamma_y$  is given in Figure 3.

- (a)  $\gamma_x - \gamma_y = 0$ :

$$\begin{aligned} D(|\psi\rangle, |\xi\rangle)^2 &= 1 - p(x)q(x) - p(y)q(y) - 2\sqrt{p(x)p(y)q(x)q(y)} \\ &= 1 - \left( \sqrt{p(x)q(x)} + \sqrt{p(y)q(y)} \right)^2 \end{aligned}$$

$$D(|\psi\rangle, |\xi\rangle) = 0 \Leftrightarrow p(x) = q(x) \quad \text{and}$$

$$D(|\psi\rangle, |\xi\rangle) = 1 \Leftrightarrow p(x) = q(y) = 0 \vee p(x) = q(y) = 1.$$

When  $p(x) = q(x)$  for all  $x \in X$  we have a perfect correlation.

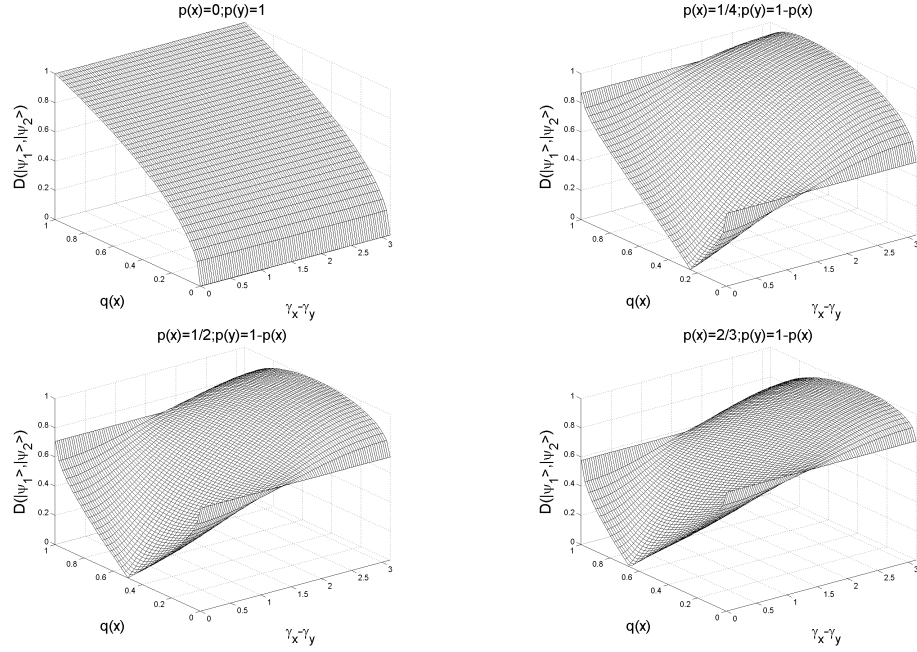


FIGURE 2. Visualization of  $D(|\psi\rangle, |\xi\rangle)$  for  $X = \{x, y\}$  and different values of  $p$ .  $q$  ranges over all density distributions and all phase differences between 0 and  $\pi$ . The range from  $-\pi$  to 0 is symmetric.

(b)  $\gamma_x - \gamma_y = \pm \frac{\pi}{3}$ :

$$\begin{aligned} D(|\psi\rangle, |\xi\rangle)^2 &= 1 - p(x)q(x) - p(y)q(y) - \sqrt{p(x)p(y)q(x)q(y)} \\ D(|\psi\rangle, |\xi\rangle) &= 0 \Leftrightarrow p(x) = q(x) = 0 \vee p(x) = q(x) = 1 \quad \text{and} \\ D(|\psi\rangle, |\xi\rangle) &= 1 \Leftrightarrow p(x) = q(y) = 0 \vee p(x) = q(y) = 1. \end{aligned}$$

(c)  $\gamma_x - \gamma_y = \pm \frac{\pi}{2}$ :

$$\begin{aligned} D(|\psi\rangle, |\xi\rangle)^2 &= 1 - p(x)q(x) - p(y)q(y) \\ D(|\psi\rangle, |\xi\rangle) &= 0 \Leftrightarrow p(x) = q(x) = 0 \vee p(x) = q(x) = 1 \quad \text{and} \\ D(|\psi\rangle, |\xi\rangle) &= 1 \Leftrightarrow p(x) = q(y) = 0 \vee p(x) = q(y) = 1. \end{aligned}$$

(d)  $\gamma_x - \gamma_y = \pm \pi$ :

$$\begin{aligned} D(|\psi\rangle, |\xi\rangle)^2 &= 1 - p(x)q(x) - p(y)q(y) + 2\sqrt{p(x)p(y)q(x)q(y)} \\ &= 1 - \left( \sqrt{p(x)q(x)} - \sqrt{p(y)q(y)} \right)^2 \\ D(|\psi\rangle, |\xi\rangle) &= 0 \Leftrightarrow p(x) = q(x) = 0 \vee p(x) = q(x) = 1 \quad \text{and} \\ D(|\psi\rangle, |\xi\rangle) &= 1 \Leftrightarrow p(x) = 1 - q(x). \end{aligned}$$

When  $p(x) = 1 - q(x)$  for all  $x \in X$  we have a perfect anti correlation.

- (3) Figure 4 shows  $D(|\psi\rangle, |\xi\rangle)$  if  $p = q$  and  $q$  is ranging over all phase differences between 0 and  $\pi$ .

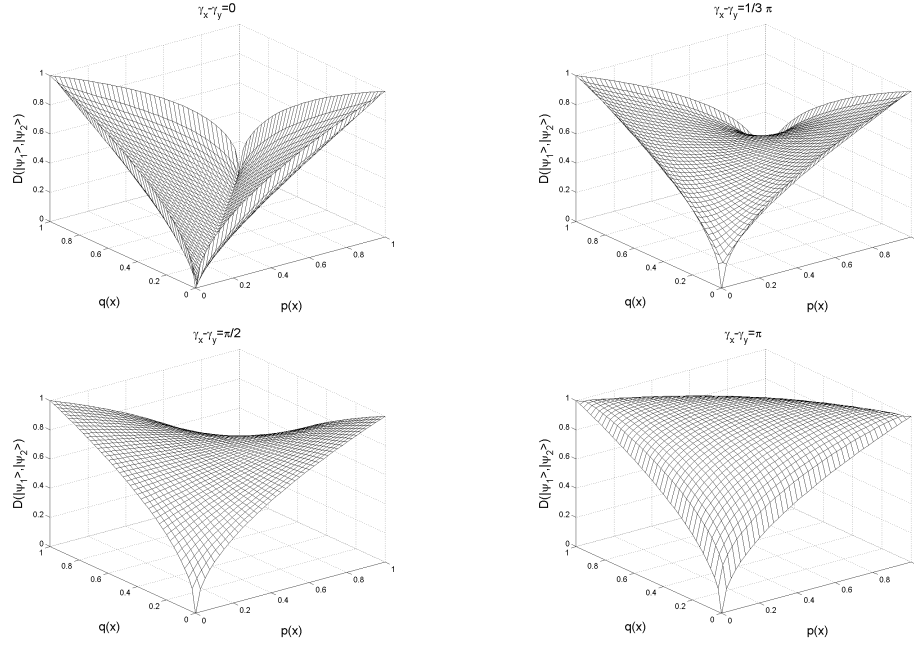


FIGURE 3. Visualization of  $D(|\psi\rangle, |\xi\rangle)$  for  $X = \{x, y\}$  and a fixed phase difference.  $p$  and  $q$  range over all possible density distributions.

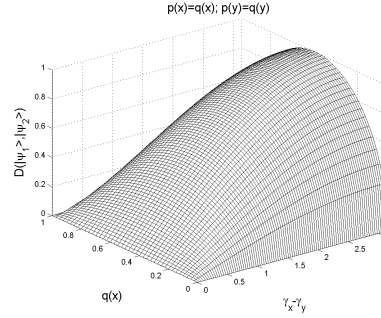


FIGURE 4. Visualization of  $D(|\psi\rangle, |\xi\rangle)$  for  $X = \{x, y\}$  and  $p = q$ . The phase difference  $\gamma_x - \gamma_y$  ranges over all phase differences between 0 and  $\pi$ . The range from  $-\pi$  to 0 is symmetric.

**4.1. Visualization of qubits, the Bloch sphere.** A very useful visualization technique that is limited to two dimensional systems is given by the so called Bloch sphere. By ignoring a physically irrelevant overall phase factor, the general state of a qubit can be written as

$$(4.1) \quad |\psi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\varphi} \sin \frac{\theta}{2} |1\rangle \quad \text{with} \quad \varphi \in [0, 2\pi), \theta \in [0, \pi].$$

If we interpret  $\theta$  and  $\varphi$  as spherical coordinates

$$(4.2) \quad \hat{r} = (\cos \varphi \sin \theta, \sin \varphi \sin \theta, \cos \theta) = (x, y, z),$$

every qubit state has a unique representation as a point on the three-dimensional unit sphere, also known as *Bloch sphere* (see Fig. 5).

The unit-vector  $\hat{r} = \hat{r}_\psi$  is called *Bloch vector* of  $|\psi\rangle$ . Bloch vectors have the property that

$$(4.3) \quad \hat{r}_\phi = -\hat{r}_\chi \iff \langle \phi | \chi \rangle = 0.$$

For density operators of a two-dimensional system a similar generalization exists. By again ignoring an overall phase, any single qubit density operator can be written as (see [18])

$$(4.4) \quad \rho = \frac{1}{2} \begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix}.$$

For mixed states, the Bloch vector  $r_\rho = (x, y, z)$  lies inside the Bloch sphere ( $\|r_\rho\| < 1$ ) and for a classical unbiased coin flip state, i.e.  $p(|0\rangle) = p(|1\rangle) = 1/2$  we get

$$(4.5) \quad \rho_c = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \Rightarrow r_{\rho_c} = (0, 0, 0).$$

Unfortunately, there is no easy generalization of the Bloch sphere for higher dimensional quantum systems.

## 5. A NEW APPROACH TO SIGNAL SEGMENTATION USING CONJUGATE INFORMATION VARIABLES

The segmentation of signals is one of the fundamental problems in the area of signal and image processing. This is especially true as many higher-level signal analysis algorithms rely heavily on the result of a low-level segmentation process. The term segmentation is mostly not well defined as it can refer to finding some objects in an image, e.g. a face, skin, coin, etc., or detecting lower-level features like edges, constant signal areas, etc.

In this chapter we suggest a new approach for a specific segmentation problem, namely the segmentation of signals into locally constant, rising or falling parts, and minima or maxima. For simplicity we assume one-dimensional signals but remark that the approach is not limited to this and can be extended to two- or higher-dimensional problems (see Fig. 6).

Let the signal be given by  $f(t)$  and let it be at least twice differentiable, that is  $f'(t)$  and  $f''(t)$  exist. In our approach the derivatives will be used as conjugate variables in order to classify local parts of functions.

We need the range of the derivatives to be within certain limits to normalize them properly. This can sometimes cause problems as we have to think about meaningful limits in a given application but usually the limits are obvious for real life signals, such as an image function where the pixel's range from zero to one, etc. The limits of the derivatives will be used as soft thresholds that control each other.

Zero crossings of the first derivative indicate extrema of a function and the sign of the second derivative distinguishes between a maximum and a minimum. The function is constant or has a saddle point if both derivatives are zero. In the presence of noise zero crossings of the first derivative becomes unstable whereas the sign of the



second derivative remains robust unless there is a saddle point. We will exploit this fact and relate it the role amplitude and phase play for qubits.

Let

$$\begin{aligned} p^\pm(t) &= \sqrt{\frac{c_2 \pm f''(t)}{2c_2}}, \\ \alpha^+(t) &= p^+(t) e^{i\pi f'(t)/c_1}, \\ \alpha^-(t) &= p^-(t), \\ &\text{with } |f'(t)| \leq c_1, |f''(t)| \leq c_2 \text{ for all } t, \end{aligned}$$

and define quantum states

$$(5.1) \quad |\psi(t)\rangle = \alpha^+(t) |z_+\rangle + \alpha^-(t) |z_-\rangle.$$

Please note that as  $|\alpha^+(t)|^2 + |\alpha^-(t)|^2 = 1$ , for all  $t$ , the  $|\psi(t)\rangle$  are well-defined qubits.  $|z_+\rangle, |z_-\rangle$  describe the positive and negative  $z$ -axis of the Bloch sphere (see Fig. 5) and  $f''(t) = \pm c_2 \iff |\psi(t)\rangle = |z_\pm\rangle$  independent of  $f'(t)$ . The positive and negative  $x$  and  $y$  axis respectively are given by

$$\begin{aligned} |x_+\rangle &= \frac{1}{\sqrt{2}}(|z_+\rangle + |z_-\rangle), & |x_-\rangle &= \frac{1}{\sqrt{2}}(|z_+\rangle - |z_-\rangle), \\ |y_+\rangle &= \frac{1}{\sqrt{2}}(|z_+\rangle + i|z_-\rangle), & |y_-\rangle &= \frac{1}{\sqrt{2}}(|z_+\rangle - i|z_-\rangle). \end{aligned}$$

$f''(t) = f'(t) = 0 \iff |\psi(t)\rangle = |x_+\rangle$  whereas  $f''(t) = 0 \wedge f'(t) = \pm c_1 \iff |\psi(t)\rangle = |x_-\rangle$ .  $f'(t)$  and  $f''(t)$  are now conjugate variables in the sense that observing the random variables  $X = \{|x_+\rangle\langle x_+|, |x_-\rangle\langle x_-|\}$  and  $Z = \{|z_+\rangle\langle z_+|, |z_-\rangle\langle z_-|\}$  the commutator<sup>3</sup> between any two projectors of  $X$  and  $Z$  becomes maximal:

$$A_\pm = |x_\pm\rangle\langle x_\pm|, B_\pm = |z_\pm\rangle\langle z_\pm| \Rightarrow \|A_\pm B_\pm - B_\pm A_\pm\| = \frac{1}{2}.$$

Next we define a set of projectors:

$$\begin{aligned} P_0 &= |z_+\rangle\langle z_+| \\ P_1 &= |z_-\rangle\langle z_-| \\ P_2 &= \frac{1}{\sqrt{2}}(|y_-\rangle\langle y_-| + |x_-\rangle\langle x_-|) = \frac{1}{\sqrt{2}}\left(P_0 - \frac{1}{\sqrt{2}}(1+i)P_1\right) \\ P_3 &= \frac{1}{\sqrt{2}}(|y_+\rangle\langle y_+| + |x_-\rangle\langle x_-|) = \frac{1}{\sqrt{2}}\left(P_0 - \frac{1}{\sqrt{2}}(1-i)P_1\right) \\ P_4 &= |x_+\rangle\langle x_+| = \frac{1}{\sqrt{2}}(P_0 + P_1) = \frac{1}{\sqrt{2}}\mathbb{I}. \end{aligned}$$

Only  $P_0$  and  $P_1$  form a random variable as they add up to the identity and therefore fulfill the completeness condition. They indicate minimum or maximum of the function.  $P_i, i = 2, \dots, 4$  would have to be complemented by  $\mathbb{I} - P_i$  in order to do so. The choice of projectors depends on the problem we want to solve.

---

<sup>3</sup>The commutator of two operators  $A, B$  is given by  $[A, B] = AB - BA$ . Two operators commute if and only if their commutator equals zero, i.e.  $AB = BA \iff [A, B] = 0$ .

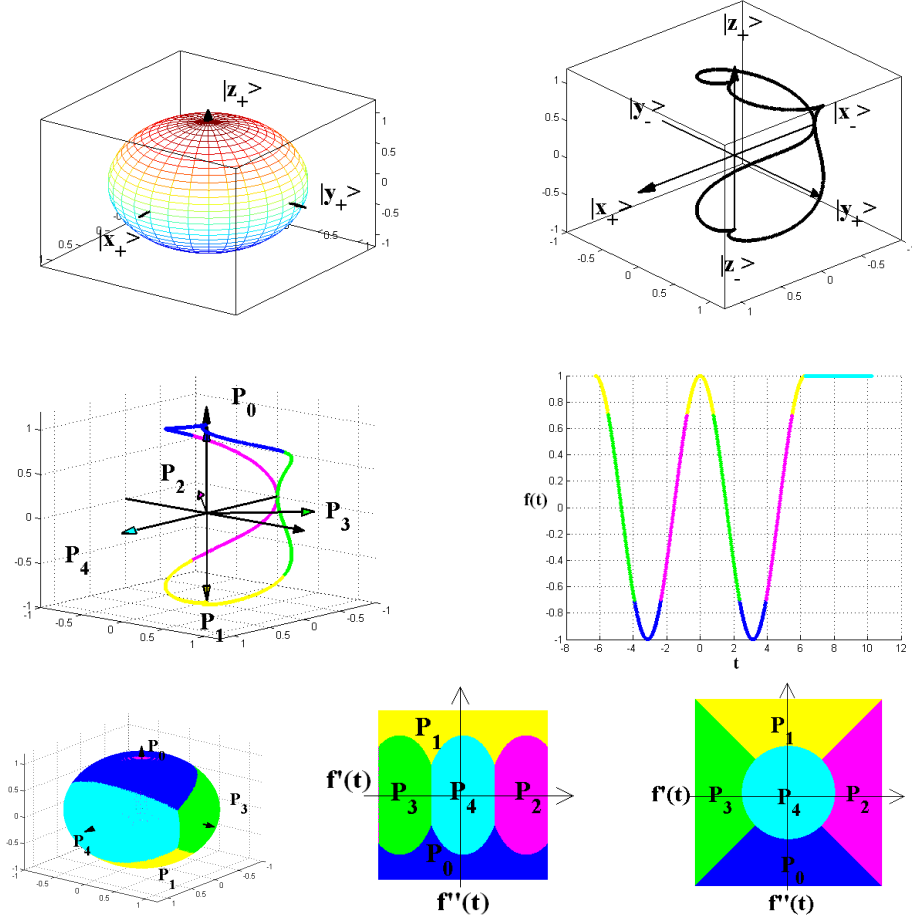


FIGURE 5. Top left: Bloch sphere representation of all qubit states. Top right: Bloch sphere representation of the qubit states  $|\psi(t)\rangle$  for  $f(t)$  as given in Equation 5.2. Middle left: Labeling result in the Bloch sphere representation. Middle right: Labeling of  $f(t)$ . Bottom left: Decision regions in the Bloch sphere representation,  $P_2$  is shining through at the north pole but is not visible from this perspective. Bottom middle: Decision regions over the joint distribution of  $f''(t)$  and  $f'(t)$ . Bottom right: Model based (see Equations 5.3) decision regions over the joint distribution of  $f''(t)$  and  $f'(t)$ . The limits of the axes are  $\pm c_1$  for  $f'(t)$  and  $\pm c_2$  for  $f''(t)$ . If  $c = c_1 = c_2$ , the radius of the inner circle,  $P_4$ , is given by  $c/2$ .

For any given  $f(t)$  we now can project the corresponding  $|\psi(t)\rangle$  and get a classification result by maximizing over all projectors

$$C(f(t)) = i \in \{0, \dots, 4\} : \langle \psi(t) | P_i | \psi(t) \rangle = \max_{j=0, \dots, 4} \langle \psi(t) | P_j | \psi(t) \rangle.$$

Alternatively, if  $X_i = \{P_i, \mathbb{I} - P_i\} = \{|\chi_{i,0}\rangle\langle\chi_{i,0}|, |\chi_{i,1}\rangle\langle\chi_{i,1}|\}$ ,  $i = 0, \dots, 4$ , and  $|\chi\rangle$  represents the eigenvector of the corresponding projector, we can minimize the distance  $D(|\chi_{i,0}\rangle, |\psi(t)\rangle)$ . The results of the classification for an analytic function

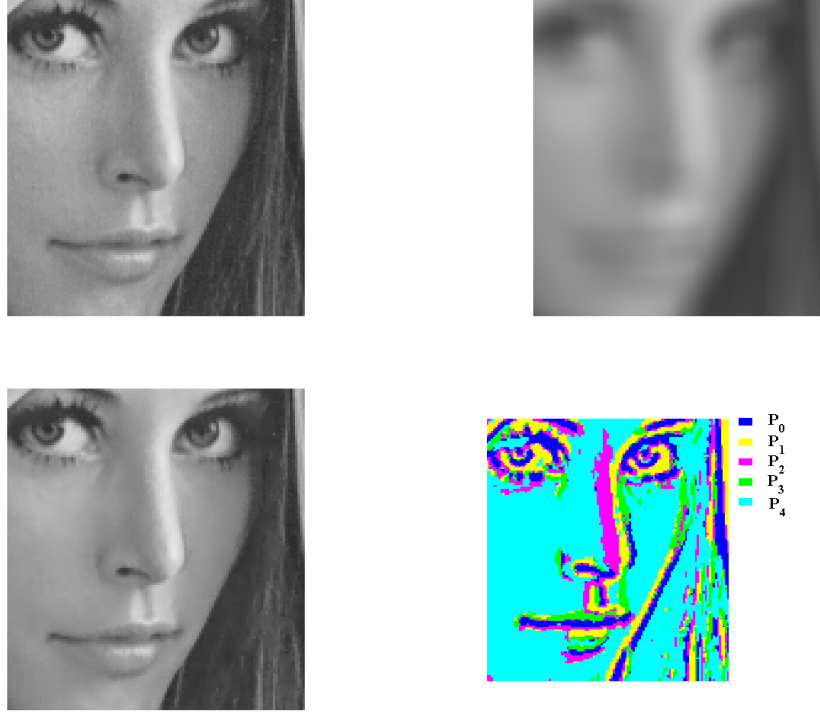


FIGURE 6. An application of the segmentation technique. The original image on the top left was smoothed on the constant part  $P_4$  of the image function as given in the bottom right picture. The result of smoothing adaptively with an average filter ( $15 \times 15$ ) is shown in the lower left picture ( $P_0, \dots, P_3$  are not filtered). The top right shows the filter result using the same filter kernel independent of the image content.

$$(5.2) \quad f(t) = \begin{cases} \cos(t), & |t| < 2\pi \\ 1, & t > 2\pi \end{cases}$$

are given in middle row of Fig. 5, whereas the bottom row (left and middle picture) shows the decision boundaries over the entire joint distribution of the derivatives as derived by the described approach. Usually we cannot access an analytic expression for  $f(t)$  but a measurement of either  $f(t)$  or  $f'(t)$  and  $f''(t)$  is available. If we measure  $f(t)$  we have to derive the derivatives numerically<sup>4</sup>. An example for the measurement of  $f'(t)$  and  $f''(t)$  is given by observing the velocity and acceleration

<sup>4</sup>Depending on the application, this might require some low pass filtering of the measurements to gain stability in the derivatives.

of an object. In general these measurements will be taken by independent devices but they are highly correlated quantities.

In either case the measurements will be subject to errors. We may choose different error models for the respective areas. In the example we gave in the introduction we could decide that the acoustic signal is mainly noise during the day time and the visual signal is not available at night time. Towards the poles of the Bloch sphere the phase term (in this case  $f'(t)$ ) becomes less important as it seamlessly transforms itself into a global phase factor. In equatorial regions it becomes predominant. A quantitative discussion of this depends on the concrete problem to be solved, especially on the statistical error model(s) to be chosen, and shall be left for a future publication.

A standard approach to derive the decision regions would be to create a model of the decision problem and then minimize some distance function over the entire joint signal distribution to this model as shown in the bottom right picture of Fig. 5 for the function given in Equation 5.2. To derive this model we have to notice the (partially) nonlinear dependency of

$$(5.3) \quad \begin{aligned} f'(t) &= -\sin(t), & f''(t) &= -\cos(t), & f'(t)^2 + f''(t)^2 &= 1, & |t| < 2\pi, \\ f'(t) &= f''(t) &= 0, & & & & t \geq 2\pi. \end{aligned}$$

In this model the decision depends on both signals over (almost) the entire joint distribution range, whereas the model derived from the conjugate variables cuts off the influence of one signal around the poles of the Bloch sphere.

## 6. DISCUSSION

The approach of modelling joint distributions of signals as conjugate variables certainly has its limitations. We remark that its strengths go along with its weaknesses. Whenever all signals are meaningful over the entire joint distribution range other approaches might be more successful<sup>5</sup>.

On the other hand in situations where we have an expectation that we can rely on some signals more than others, depending on some given (or derived) parameters, our approach will be valuable as this dependency may be modelled directly. It is noteworthy that the role of the two signals used for the encoding might be quite different. In the bird example we may classify the two species, let's call them *Happy* and *Unhappy*, using the acoustic signal alone at times but even when this signal becomes mixed we still can rely on its presence and then take the optical signal into account. The converse is not true as we can not speak about a colour at night. Therefore, the optical sensor (used as phase information) might give arbitrary measurement results (including blue or green!) at night time. This will not influence the classification result as the measurement of the conjugate acoustic signal is mapped to one of the polar regions at night time. In lucky circumstances, e.g. when by chance we have no background noise during the day or the moon provides enough light at night, the joint observation of chirp and colour does not change the classification result either, as a *Happy* bird does not become more *Happy* if we know both features. When the noise level is slowly increasing (at sunrise or sunset)

---

<sup>5</sup>In the Hilbert space setting used in this paper we need a higher dimensional system, e.g. a two qubit system. Of course, we may add conjugate information to each of those systems, if it is available to us.

seamlessly both signals contribute to the classification and allow a discrimination of the two classes.

Just to give a flavour of the results to expect, Fig. 6 shows the application of an adaptive filter on image functions that is based on the segmentation approach. At the top left the original image (used as the signal  $f(t)$ ) is given which is to be smoothed to remove noise introduced by the sensor (camera, scanner, etc.). If we assume uniform noise, we can remove this by averaging over small parts of the image function. The application of an average filter independent of the image structure will remove this noise but unfortunately it will remove parts of the image content at the same time as can be seen in the top right picture of Fig. 6. If we classify each point of the image according to the approach described in Section 5, we can apply the averaging process only to image areas that are labelled constant, i.e.  $P_4$  (Fig. 6 bottom right). This results in an image that is fairly smooth in these constant image areas but untouched everywhere else (Fig. 6 bottom left). In the other areas we could apply different filter kernels (and different error models) if we wish to do so.

## 7. CONCLUSION AND FUTURE WORK

In this paper we have described a method for modelling joint distributions of signals as conjugate variables in finite-dimensional complex Hilbert spaces. We have derived a distance function based on the transition probability of quantum states that is a metric on pure states and satisfies the triangle inequality for mixed states, and we have related it to some distance functions between probability distributions. The analysis of other distance measures on quantum states, especially the so called trace distance which is closely related to the variational distance of probability distributions, will be a future task.

We believe that the mathematical concept of quantum information theory offers a fertile resource in many areas of information processing. In this article we mainly focused on non-decomposed principle systems, but it would be very interesting to analyze the decomposition of higher-dimensional systems which remains a future task. We have not touched on the field of quantum operations, i.e. unitary transformations of quantum states which again offers a rich potential. Additionally generalized measurement operators (positive operator valued measurements (POVM)) offer a valuable future research direction. The use of Wigner functions provides another promising research area. Wigner functions relate the probability distributions of conjugate variables to each other as they describe this joint (real valued) distribution. Moreover they relate the spatial to the frequency domain, which may have further impact.

## REFERENCES

1. F. J. Aherne, N. A. Thacker, and P.I. Rockett, *The Bhattacharyya metric as an absolute similarity measure for frequency coded data*, Kybernetika **32** (1997), no. 4, 001–007.
2. Robert B. Ash, *Information theory*, Dover Publication, Inc., New York, 1965.
3. O. E. Barndorff-Nielsen, R. D. Gill, and P. E. Jupp, *On quantum statistical inference*, J. Roy. Statist. Soc. (B) **65** (2003), 775–816.
4. D.C. Brody and L.P. Hughston, *Quantum noise and stochastic reduction*, Journal of Physics A, MATH GEN **39** (2006), 833–876, ISSN: 0305-4470.
5. M. W. Coffey, *Properties and possibilities of quantum shapelets*, Journal of Physics A **39** (2006), 877–887.

6. D. Comaniciu, V. Ramesh, and P. Meer, *Kernel-based object tracking*, IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003), no. 5, 564–577.
7. T.M. Cover and J.A. Thomas, *Elements of information theory*, 1st ed., John Wiley & Sons, New York, 1991.
8. Yonina C. Eldar and Alan V. Oppenheim, *Quantum signal processing*, IEEE Signal Processing Magazine (2002).
9. C. A. Fuchs, *Distinguishability and accessible information in quantum theory*, Online preprint quant-ph/9601020, 1996.
10. Robert G. Gallager, *Information theory and reliable communication*, John Wiley & Sons, New York, 1968.
11. Lov Grover and Terry Rudolph, *Creating superpositions that correspond to efficiently integrable probability distributions*, <http://arxiv.org/abs/quant-ph/0208112>, 2002.
12. Josef Gruska, *Quantum computing*, McGraw-Hill, London, 1999.
13. C. W. Helstrom, *Quantum detection and estimation theory*, Academic Press, New York, 1976.
14. Ulrike Herzog and János A. Bergou, *Distinguishing mixed quantum states: Minimum-error discrimination versus optimum unambiguous discrimination*, Phys. Rev. A **70** (2004), no. 2, 022302.
15. A.S. Holevo, *Probabilistic and statistical aspect of quantum theory*, North Holland, P.C., 1982.
16. Solomon Kullback, *Information theory and statistics*, Dover Publication, Inc., New York, 1997.
17. J. Lin, *Divergence measures based on the Shannon entropy*, IEEE Transactions on Information Theory **37** (1991), no. 1, 145–151.
18. Zdzislaw Meglicki, *Introduction to quantum computing (m743)*, <http://beige.ucs.indiana.edu/M743/>, April 2005.
19. Michael A. Nielsen and Isaac L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, 2000.
20. Michael Nölle and Bernhard Ömer, *Representation of cyclic colour spaces within quantum space*, Proceedings of the 8<sup>th</sup> International Conference on Digital Image Computing - Techniques and Applications (DICTA'05) (2005).
21. Michael Nölle, Bernhard Ömer, and Martin Suda, *Quantum information algorithms - new solutions for known problems*, e & i Elektrotechnik und Informationstechnik, Springer Wien **124** (2007), no. 5.
22. Puzicha, J., Rubner, Y., Tomasi, C., and Buhmann, J. M., *Empirical Evaluation of Dissimilarity Measures for Color and Texture*, Proc. of the International Conference on Computer Vision (ICCV'99), 1999, pp. 1165–1173.
23. M. Zaharieva, R. Huber-Mörk, M. Nölle, and M. Kampel, *On ancient coin classification*, International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST, 2007<sup>8th</sup> (2007).